From Matching to Generation: A Survey on Generative Information Retrieval

Year: 2024 | Citations: 123 | Authors: Xiaoxi Li, Jiajie Jin, Yujia Zhou, Yuyao Zhang, Peitian Zhang

---

Abstract

Information Retrieval (IR) systems are crucial tools for users to access information, which have long been dominated by traditional methods relying on similarity matching. With the advancement of pre-trained language models, Generative Information Retrieval (GenIR) emerges as a novel paradigm, attracting increasing attention. Based on the form of information provided to users, current research in GenIR can be categorized into two aspects: (1) Generative Retrieval (GR) leverages the generative model's parameters for memorizing documents, enabling retrieval by directly generating relevant document identifiers without explicit indexing. (2) Reliable Response Generation employs language models to directly generate information users seek, breaking the limitations of traditional IR in terms of document granularity and relevance matching while offering flexibility, efficiency, and creativity to meet practical needs. This article aims to systematically review the latest research progress in GenIR. We will summarize the advancements in GR regarding model training and structure, document identifier, incremental learning, and so on, as well as progress in reliable response generation in aspects of internal knowledge memorization, external knowledge augmentation, and so on. We also review the evaluation, challenges, and future developments in GenIR systems. This review aims to offer a comprehensive reference for researchers, encouraging further development in the GenIR field (Github Repository: https://github.com/RUC-NLPIR/GenIR-Survey).